UNIVERSITY OF LONDON

GOLDSMITHS' COLLEGE

B. Sc. Examination 2003

STATISTICS

ST52007A (ST218) Principles and Methods of Statistics

Duration: 2 hours 15 minutes

Date and time:

Answer ALL FIVE questions from Section A, and ONE question from Section B
There are 80 marks available on this paper.
A Formula Sheet is attached to the end of this examination paper.
Electronic calculators may be used. The make and model should be specified on the script. The calculator must not be programmed prior to the examination. Calculators which display graphics, text or algebraic equations are not allowed.
NOTE: Full details of all calculations are to be shown; pre-programmed statistical tests and procedures on a calculator, apart from mean and standard deviation, must not be used.
WHITE, YEATS & SKIPWORTH: Tables for Statisticians to be provided.

THIS EXAMINATION PAPER MUST NOT BE REMOVED FROM THE EXAMINATION ROOM

Part A: Answer ALL questions

Question 1 The random variable X has a Gamma $Ga(\alpha, \lambda)$ distribution where $\alpha > 0, \lambda > 0$.

(a) Given that the pdf of a $Ga(r + \alpha, \lambda)$ distribution is proper, show that

$$E[X^r] \equiv \mu'_r = \frac{\Gamma(r+\alpha)}{\Gamma(\alpha)\lambda^r}, \quad (r > -\alpha). \qquad [5]$$

(b) Hence show that the variance of X is given by $\sigma^2 = \alpha/\lambda^2$ and that the standardised coefficient of skewness $\gamma_1 \equiv E[(X-\mu)^3]/\sigma^3 = 2/\sqrt{\alpha}$.
(HINT: use the identity $E[(X-\mu)^3] = \mu'_3 - 3\mu\mu'_2 + 2\mu^3$.) [6]

Question 2 A random sample $x_1, \ldots, x_n$ with mean $\bar{x}$ is observed from the $Ex(\lambda)$ distribution.

(a) Show that the likelihood function is given by

$$L(\lambda) = \lambda^n e^{-n\lambda\bar{x}}. \qquad [2]$$

(b) Hence show that the MLE of λ is $\hat{\lambda} = 1/\bar{x}$, verifying that the likelihood is indeed maximised. [4]

(c) Hence find the information function $I(\lambda)$, and write down a large sample 99% confidence interval for λ in terms of n and $\bar{x}$, stating the theorem you are using. [5]

Question 3 (a) Show that the sampling distribution of the mean of a random sample from $Ex(\lambda)$ is $Ga(n, n\lambda)$. [4]

(b) Hence using the results of Question 1, show that the bias of the MLE $\hat{\lambda}$ in Question 2 is $\lambda/(n-1)$, [2]

(c) Further show that

$$\text{Var}[\hat{\lambda}] = \frac{\lambda^2 n^2}{(n-1)^2(n-2)}. \qquad [5]$$

Question 4 A random sample $x_1, \ldots, x_n$ is taken from a Negative Binomial $NB(k, p)$ distribution, with k known. A Bayesian analysing these data chooses a Beta prior distribution for p with parameters α and β.

(a) Show that the posterior distribution for p is also Beta, and find its parameters. [4]

(b) Hence show that the posterior mean is a weighted linear combination of the prior mean and MLE $\hat{p} = \bar{x}/(\bar{x} + k)$, and find the weight attached to the MLE. [4]

(c) Show that this weight tends to 1 not only as the sample size $n \to \infty$, but also as k increases, and as the prior parameters both tend to zero. [3]

Question 5 A *logarithmic* distribution with parameter θ with p.d.f. given by

$$f(x \mid \theta) = a(\theta)\frac{\theta^x}{x}, \qquad x = 1, 2, \ldots,$$

where $a(\theta) = -1/\log(1-\theta)$, $0 < \theta < 1$, has mean $\frac{a(\theta)\theta}{1-\theta}$.

(a) Show that the loglikelihood from a random sample $x_1, \ldots, x_n$ is, up to an additive constant,

$$l(\theta) = n \log a(\theta) + n\bar{x} \log \theta. \quad [3]$$

(b) Hence show that the MLE $\hat{\theta}$ solves the method of moments equation

$$\bar{x} = \frac{a(\theta)\theta}{1-\theta}. \quad [3]$$

(c) Hence also show that the information function $I(\theta)$ is given by

$$I(\theta) = \frac{na(\theta)[1 + \theta a(\theta)]}{\theta(1-\theta)^2}. \qquad [5]$$

Part B: Answer ONE question only

Question 6 The joint p.d.f. $f(x,y)$ of two random variables X, Y is given by

$$f(x,y) = 1 + \alpha(1-2x)(1-2y), \quad 0 < x < 1, \quad 0 < y < 1, \quad \text{zero otherwise,}$$

where $-1 \leq \alpha \leq 1$.

(a) Show that the joint c.d.f. $F(x,y) = \Pr[X \leq x,\ Y \leq y]$ is given by

$$F(x,y) = xy[1 + \alpha(1-x)(1-y)], \quad 0 \leq x \leq 1, \quad 0 \leq y \leq 1 \qquad [8]$$

(b) Hence find the marginal c.d.f.s of both X and Y, showing they are both those of a $U(0,1)$ random variable. [4]

(c) Show that the correlation ρ of X and Y is given by $\rho = \alpha/3$. [8]

(d) Find also the conditional p.d.f. of Y given $X = x$ and show that its mean is $\rho x + \frac{1}{2}(1-\rho)$. [5]

Question 7 Accidents at a busy junction are thought to follow a Poisson process so that the number of accidents X_i in a time period of length t_i (known) has a Poisson distribution with mean $\mu_i = \lambda t_i$ $(i = 1, \ldots, n)$, where $\lambda > 0$ is an unknown parameter. The parameter λ represents the accident rate to be estimated from the n non-overlapping periods of observation.

(a) Show that the likelihood function $L(\lambda)$ is proportional to $e^{-\lambda \sum_{i=1}^n t_i} \lambda^{\sum_{i=1}^n x_i}$. [4]

(b) Sketch the likelihood function. [2]

(c) Show that the maximum likelihood estimate (MLE) $\hat{\lambda}$ satisfies

$$\hat{\lambda} = \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n t_i}. \qquad [3]$$

(d) Show that an approximate 95% confidence interval for λ is given by the end-points $(\sum_{i=1}^n x_i \pm 1.96\sqrt{\sum_{i=1}^n x_i}) / \sum_{i=1}^n t_i$. [6]

(e) Find additionally the maximised value of the loglikelihood. [3]

(f) The *saturated model* postulates a different accident rate λ_i, say for the i^{th} period $(i = 1, \ldots, n)$ and hence the MLEs are $\dot{\lambda}_i = x_i/t_i$, $i = 1, \ldots, n$. Find the maximised value of its loglikelihood and show that the difference is

$$\sum_{i=1}^n x_i \left[\log\left(\frac{x_i}{\sum_{i=1}^n x_i}\right) - \log\left(\frac{t_i}{\sum_{i=1}^n t_i}\right)\right]. \qquad [8]$$